Performance guarantees for a variational ”multi-space” decoder

Cédric Herzet, Mamadou Diallo

Noname manuscript No.
(will be inserted by the editor)

Performance guarantees for a variational "multi-space" decoder

C. Herzet[1] · M. Diallo[1]

Received: date / Accepted: date

Abstract Model-order reduction methods tackle the following general approximation problem: find an "easily-computable" but accurate approximation $\hat{\boldsymbol{h}}$ of some target solution $\boldsymbol{h}^\star$. In order to achieve this goal, standard methodologies combine two main ingredients: *i)* a set of partial observations of $\boldsymbol{h}^\star$; *ii)* some "simple" prior model on the set of target solutions. The most common prior model encountered in the literature assume that the target solution $\boldsymbol{h}^\star$ is "close" to some low-dimensional subspace. Recently, triggered by the work by Binev *et al.* [3], several contributions have shown that refined prior models (based on a *set* of embedded approximation subspaces) may lead to enhanced approximation performance. In this paper, we focus on a particular decoder exploiting such a "multi-space" information and evaluating $\hat{\boldsymbol{h}}$ as the solution of a constrained optimization problem. To date, no theoretical results have been derived to support the good empirical performance of this decoder. The goal of the present paper is to fill this gap. More specifically, we provide a mathematical characterization of the approximation performance achievable by this variational "multi-space" decoder and emphasize that, in some specific setups, it has provably better recovery guarantees than its standard "single-space" counterpart. We also discuss the similarities and differences between this decoder and the one proposed in [3].

Keywords Model-order reduction · Multi-space prior information · Performance guarantees

The authors thank the "Agence nationale de la recherche" for its financial support through the Geronimo project (ANR-13-JS03-0002).

C. Herzet
Tel.: +33-2-99847350
E-mail: cedric.herzet@inria.fr

[1] INRIA, Centre Rennes Bretagne-Atlantique, Campus de Beaulieu, 35000 Rennes, France

1 Introduction

Many approximation methods encountered in the domain of model-order reduction rely on the following ingredients:

1) some *"prior" knowledge* about the location of the target solution $\boldsymbol{h}^\star$ in the ambient space $\mathcal{H}$.[1] This prior information typically takes the form of a subset $\mathcal{M}_{\text{prior}} \subset \mathcal{H}$ such that $\boldsymbol{h}^\star \in \mathcal{M}_{\text{prior}}$.
2) a set of *partial observations* of $\boldsymbol{h}^\star$, say $\{y_j\}_{j=1}^m$. In this paper, we assume that $y_j = \langle \boldsymbol{w}_j, \boldsymbol{h}^\star \rangle$ where $\{\boldsymbol{w}_j\}_{j=1}^m$ is a basis of some m-dimensional subspace W_m.

Examples of problems exploiting these two ingredients include "projection-based" reduction of linear parametrized partial differential equations (PPDE) [11], some approximation methods for non-linear operators (*e.g.*, GEIM [9], DEIM [4], Gappy POD [5]) or data-assimilation schemes with reduced-order models [3].

Both $\mathcal{M}_{\text{prior}}$ and $\{y_j\}_{j=1}^m$ provide some information about the location of $\boldsymbol{h}^\star$ in $\mathcal{H}$. In particular, in the case of linear measurements, observing $\{y_j\}_{j=1}^m$ indicates that $\boldsymbol{h}^\star$ must belong to some m-dimensional affine subspace of $\mathcal{H}$:

$$\begin{aligned} \boldsymbol{h}^\star \in \mathcal{P}_{\mathbf{y}} &\triangleq \{\boldsymbol{h} \in \mathcal{H} : \langle \boldsymbol{w}_j, \boldsymbol{h} \rangle = y_j \text{ for } j = 1 \dots m\} \\ &= \big\{\boldsymbol{h}^\star + \boldsymbol{h} : \boldsymbol{h} \in W_m^\perp\big\}, \end{aligned} \tag{1}$$

where $W_m^\perp$ is the orthogonal complement of W_m in $\mathcal{H}$. The goal of an approximation method then consists in combining effectively (in some sense) these two sources of information to provide a good approximation $\hat{\boldsymbol{h}}$ of $\boldsymbol{h}^\star$.

The choice of $\mathcal{M}_{\text{prior}}$ and W_m usually plays a crucial role in the performance of the approximation methods since they determine a trade-off between accuracy and complexity. A standard option consists in choosing $\mathcal{M}_{\text{prior}}$ as:

$$\mathcal{M}_{\text{prior}} = \{\boldsymbol{h} \in \mathcal{H} : \operatorname{dist}(\boldsymbol{h}, V_n) \le \hat{\epsilon}_n\} \tag{2}$$

where V_n is an n-dimensional subspace, $\operatorname{dist}(\boldsymbol{h}, V_n) \triangleq \min_{\boldsymbol{h}' \in V_n} \|\boldsymbol{h} - \boldsymbol{h}'\|$ and $\hat{\epsilon}_n \ge 0$. Prior model (2) indicates that the sought solution $\boldsymbol{h}^\star$ is close (up to some error bounded by $\hat{\epsilon}_n$) to an n-dimensional subspace V_n. This simple model has been exploited in most contributions of the literature, see *e.g.*, [4, 5, 9, 11].

Recently, triggered by the work by Binev *et al.* [3], more refined definitions of $\mathcal{M}_{\text{prior}}$ have been considered in [1, 6–8]. In these works, $\mathcal{M}_{\text{prior}}$ is built from a sequence of embedded approximation subspaces, that is

$$\mathcal{M}_{\text{prior}} = \cap_{k=0}^n \{\boldsymbol{h} \in \mathcal{H} : \operatorname{dist}(\boldsymbol{h}, V_k) \le \hat{\epsilon}_k\} \tag{3}$$

where $\hat{\epsilon}_k \ge 0$, $\dim(V_k) = k$ and

$$V_0 \subset V_1 \subset \ldots \subset V_n. \tag{4}$$

[1] In this paper, we assume that $\mathcal{H}$ is a Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and induced norm $\|\cdot\|$.

Prior (3) is usually referred to as "multi-space" prior. It indicates that $\boldsymbol{h}^\star$ is located at some bounded distances of a sequence of embedded subspaces $\{V_k\}_{k=0}^n$. This kind of information may for example stem from prior physical knowledge about the process under study or come as a by-product of the procedures computing the approximation subspaces, see *e.g.*, [11].

Two different decoders exploiting multi-space information were proposed in [3] and [1, 6–8]. In [3], the authors suggested to approximate $\boldsymbol{h}^\star$ by some feasible point of the set $\mathcal{M}_{\text{prior}} \cap \mathcal{P}_{\mathbf{y}}$. They showed theoretically that this particular decoder may outperform standard single-space procedures in specific settings. In [1, 6–8] a different approach was considered: an approximation $\hat{\boldsymbol{h}}$ of $\boldsymbol{h}^\star$ was computed as the solution of an optimization problem with constraints $\hat{\boldsymbol{h}} \in \mathcal{M}_{\text{prior}}$ and $\hat{\boldsymbol{h}} \in V_n$.[2] The authors showed empirically that this procedure may have superior approximation performance than single-space approaches in different applicative domains. The goal of the present work is to give some theoretical ground to these observations. More specifically, we derive theoretical guarantees relating the approximation performance of the decoder considered in [1, 6–8] to the distances between the target solution $\boldsymbol{h}^\star$ and the approximation subspaces $\{V_k\}_{k=0}^n$.[3] We show morerover in several examples that, for particular choices of subspaces $\{V_k\}_{k=0}^n$ and W_m, this decoder can lead to superior approximation guarantees than its standard counterpart (2).

The rest of this paper is organized as follows. In Section 2, we present several applicative domains where the general framework described in this introduction may apply. In Section 3, we recall the standard reconstruction guarantees holding in the single-space setup, *i.e.*, when the prior information exploited by the approximation procedure only involves *one* n-dimensional subspace V_n. Section 4 is dedicated to the multi-space setup. In Section 4.1, we give a precise definition of the multi-space decoder considered in [1, 6–8] and provide some theoretical guarantees on its performance (see Theorem 2). In Section 4.2, we particularize this result to different setups and show that the multi-space decoder proposed in [1, 6–8] has provably more favorable performance than the single-space approach in these cases. We also provide a graphical interpretation of the geometry underlying the single and multi-space problems, which leads to some further insights into the scenarios where multi-space priors may be of interest. Finally, in Section 4.3 we discuss the similarities and differences between the multi-space decoders proposed in [3] and [1, 6–8]. In particular we show that when $\{\boldsymbol{w}_j\}_{j=1}^m$ is an orthonormal basis, the theoretical guarantees derived in this paper are slightly more favorable than those in [3] but that no general ordering of the decoder performances is possible. The proof of our main result is finally detailed in Section 5.

---

[2] We give a precise formulation of the decoders considered in [3] and [1, 6–8] in Section 4 of this paper.

[3] We note that the main result of this paper (see Theorem 2 in Section 4) was presented in the scientific conferences Enumath'17 [7] and Morepas'18 [8] but not published elsewhere with a complete description of the proofs.

2 Standard approximation problems

In this section, we present several particularizations of the general approximation problem introduced in Section 1. This allows us to replace the different contributions of the literature dealing with multi-space priors in their context and motivate our theoretical results in Section 4 from an applicative perspective.

2.1 Reduced-order modeling

In the context of reduced-order modeling of PPDEs, see *e.g.*, [11], the goal consists in approximating the elements of a set of solutions:

$$\mathcal{M} = \{\boldsymbol{h}^\star \in \mathcal{H} : \mathrm{PDE}(\boldsymbol{h}^\star, \theta) = 0 \text{ for some } \theta \in \Theta\} \tag{5}$$

where Θ is some set of parameters and $\mathrm{PDE}(\boldsymbol{h}^\star, \theta) = 0$ is an abstract notation for the PPDE. In the case of elliptic PPDEs, the weak formulation of the differential problem reads:

$$\text{Find } \boldsymbol{h}^\star \in \mathcal{H} \text{ such that } \; a_\theta(\boldsymbol{h}^\star, \boldsymbol{h}) = b_\theta(\boldsymbol{h}) \quad \forall \boldsymbol{h} \in \mathcal{H}, \tag{6}$$

where $a_\theta : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ and $b_\theta : \mathcal{H} \to \mathbb{R}$ are some problem-specific bilinear and linear operators depending on θ.

If $\mathcal{H}$ is infinite (or very high) dimensional, solving (6) may be computationally expensive. In order to alleviate this problem, model reduction combines a subset of the constraints defined in (6) with some prior information on $\boldsymbol{h}^\star$ to compute an approximated solution $\hat{\boldsymbol{h}}$. For example, in the context of single-space prior information (2), a typical approach consists in imposing "$\hat{\boldsymbol{h}} \in V_n$" where V_n is some proper n-dimensional subspace. Moreover, the subset of constraints in (6) used in the reduced model is usually defined as

$$\text{"}\hat{\boldsymbol{h}} \text{ must be such that } a_\theta(\hat{\boldsymbol{h}}, \boldsymbol{z}_j) = b_\theta(\boldsymbol{z}_j) \quad \forall 1 \leq j \leq m\text{"}, \tag{7}$$

where $\{\boldsymbol{z}_j\}_{j=1}^m$ is a family of linearly-independent elements of $\mathcal{H}$.

We note that (7) is equivalent to imposing a constraint of the form (1). Indeed, letting $\boldsymbol{w}_j \in \mathcal{H}$ be the Riesz representer of the linear form $a_\theta(\cdot, \boldsymbol{z}_j)$, we have that

$$\langle \boldsymbol{w}_j, \boldsymbol{h}^\star \rangle = a_\theta(\boldsymbol{h}^\star, \boldsymbol{z}_j) = b_\theta(\boldsymbol{z}_j), \tag{8}$$

where the last equality follows from (6). Hence, defining $y_j \triangleq b_\theta(\boldsymbol{z}_j)$ and $W_m \triangleq \operatorname{span}\left(\{\boldsymbol{w}_j\}_{j=1}^m\right)$, it is easy to see that (7) is tantamount to enforcing "$\hat{\boldsymbol{h}} \in \mathcal{P}_{\mathbf{y}}$". We note that in this particular example, the definition of $\boldsymbol{w}_j$ depends on θ and thus varies for each problem instance.

Another example of approximation problem closely-connected to model reduction of PPDEs is the time-integration procedure considered in [6]. In this contribution, the use of a semi-implicit integration scheme requires the resolution of a problem similar to (6) at each time step. In this particular case, the bilinear operator appearing in the approximation problem (and thus the definition of the subspace W_m) depends on the solution of the integration scheme at the previous time step. The authors exploit a particular multi-space procedure (see Section 4.1) to find a good approximated solution of this problem.

2.2 Generalized empirical interpolation and data assimilation

The "Generalized Empirical Interpolation" and "Data Assimilation" problems, considered respectively in [1, 9] and [3], are closely connected to the model-reduction setup described above. In these frameworks, the set of target solutions $\mathcal{M}$ takes the form of a compact subset of $\mathcal{H}$ and the goal is to recover $\boldsymbol{h}^\star \in \mathcal{M}$ (up to some accuracy) from a set of m linear measurements of $\boldsymbol{h}^\star$, say $\{\langle \boldsymbol{w}_j, \boldsymbol{h}^\star\rangle\}_{j=1}^m$.

As a typical example of such a setup, one may mention (see *e.g.*, [9]) the case where $\mathcal{M}$ corresponds to a subset of the square-integrable functions on a bounded domain $\Omega \subset \mathbb{R}^d$ (*e.g.*, defining the response of a physical system to different operating conditions), and $\{\boldsymbol{w}_j\}_{j=1}^m$ characterizes the behavior of some sensing devices probing the state of the physical system at different locations, *e.g.*,[4]

$$\begin{aligned} \boldsymbol{w}_j : \Omega &\to \mathbb{R} \\ \mathbf{x} &\mapsto \exp\left(-\tfrac{\|\mathbf{x}-\mathbf{m}_j\|_2^2}{2\sigma^2}\right) \end{aligned} \tag{9}$$

where $\mathbf{m}_j \in \Omega$ and $\sigma^2 > 0$ are some parameters. Going back to the generic approximating problem introduced in Section 1, we see that the "observation" subspace W_m is here trivially given by $W_m = \mathrm{span}(\{\boldsymbol{w}_j\}_{j=1}^m)$.

Different priors have been considered in the literature to complement the collected observations. The standard GEIM is based on single-space information, where the estimate $\hat{\boldsymbol{h}}$ is contrained to belong some n-dimensional subspace V_n. Recently, GEIM has been reformulated in the context of multi-space information (3) in [1]. The decoder considered in this contribution is presented in Section 4.1. In [3] both single-space and multi-space prior information were considered in the context of data assimilation. The multi-space proposed in this contribution is different from the one considered in the present paper and its definition is recalled in Section 4.3.

3 The single-space approximation

In this section, we review some of the guarantees of performance holding in the single-space setup, that is when prior information of the form (2) is available.

[4] In this example, $\boldsymbol{w}_j$ models the fact that the state of the system can be sensed (with some precision σ^2) around $\mathbf{m}_j \in \Omega$.

The material presented in this section is not novel but is intended to support our discussion in the next section.

When dealing with single-space approximations, a common option to compute an approximation of $\boldsymbol{h}^\star$ consists in solving the following problem:

$$\text{Find } \hat{\boldsymbol{h}}_{\mathrm{SS}} \in V_n \text{ such that } \left\langle \boldsymbol{w}_j, \hat{\boldsymbol{h}}_{\mathrm{SS}} \right\rangle = y_j \text{ for all } j = 1 \ldots m. \tag{10}$$

In a nutshell, (10) is equivalent to finding an element of V_n satisfying the observation constraints. Problem (10) is for example the standard approach used in reduced modeling or GEIM. The setup of most interest in this case is $n = m$, since a unique solution to (10) then exists under mild conditions (see Theorem 1 below).

We note that, in the general case $\boldsymbol{h}^\star \notin V_n$ and therefore $\hat{\boldsymbol{h}}_{\mathrm{SS}} \neq \boldsymbol{h}^\star$. Nevertheless, depending on the problem at stake, $\hat{\boldsymbol{h}}_{\mathrm{SS}}$ may "close" to $\boldsymbol{h}^\star$ in some situations. In fact, the quality of the approximation obtained from the single-space problem (10) can be characterized as a function of the subspaces W_m and V_n involved in this problem. More specifically, typical recovery results in the single-space setup take the form

$$\left\| \hat{\boldsymbol{h}} - \boldsymbol{h}^\star \right\| \leq C(W_m, V_n) \operatorname{dist}(\boldsymbol{h}^\star, V_n) \tag{11}$$

where $C(W_m, V_n)$ is some constant depending on W_m and V_n. For example, in the context of Hilbert spaces the following result holds:

Theorem 1 *Let $m = n$ and*

$$\sigma_{\min} \triangleq \inf_{\boldsymbol{h} \in V_n} \sup_{\boldsymbol{h}' \in W_m} \frac{\langle \boldsymbol{h}', \boldsymbol{h} \rangle}{\|\boldsymbol{h}'\| \|\boldsymbol{h}\|}. \tag{12}$$

If $\sigma_{\min} > 0$, (10) admits a unique solution $\hat{\boldsymbol{h}}_{\mathrm{SS}}$ and

$$\left\| \boldsymbol{h}^\star - \hat{\boldsymbol{h}}_{\mathrm{SS}} \right\| \leq \sigma_{\min}^{-1} \operatorname{dist}(\boldsymbol{h}^\star, V_n). \tag{13}$$

Moreover, (13) is tight in the following sense: $\forall \epsilon_n > 0$, $\exists \boldsymbol{h}^\star \in \mathcal{H}$ such that $\operatorname{dist}(\boldsymbol{h}^\star, V_n) = \epsilon_n$ and

$$\left\| \boldsymbol{h}^\star - \hat{\boldsymbol{h}}_{\mathrm{SS}} \right\| = \sigma_{\min}^{-1} \epsilon_n. \tag{14}$$

A proof of this result can be found in [9, Theorem 2.4] for the first part and [3, Theorem 2.9] for the second. We note that (13) is only one particular example of guarantee valid for single-space approximations in Hilbert spaces. Different guarantees exist in other setups. For example, in Banach spaces (11) holds with a slightly larger constant $C(W_n, V_n) = 1 + \sigma_{\min}^{-1}$, see [9, Theorem 1.3]. In the context of reduced-order model (cf. Section 2.1), similar results exist where the constant $C(W_n, V_n)$ is related to the "inf-sup stability" of the operator characterizing the PPDE, see *e.g.*, [2]. Other works also characterize the link between the worst-case approximation error and the Kolmogorov width of a set of target solutions, see *e.g.*, [10]. In what follows, we will restrict our attention to (13), since it is valid in the context of Hilbert spaces considered in this paper, and is more amenable to comparisons with our main result stated in Theorem 2.

Before concluding this section, let us mention that the authors in [3, Theorem 2.9] showed that the approximation computed in (10) is in fact the “worst-case” optimal solution when single-space prior information (2) is available. More specifically, if $m = n$ and $\sigma_{\min} > 0$, (10) can be rewritten as

$$\hat{\boldsymbol{h}}_{\mathrm{SS}} = \operatorname*{arg\,min}_{\boldsymbol{h}\in\mathcal{H}} \sup_{\boldsymbol{h}'\in\mathcal{M}_{\mathrm{prior}}\cap\mathcal{P}_{\mathbf{y}}} \left\|\boldsymbol{h}-\boldsymbol{h}'\right\|, \tag{15}$$

where $\mathcal{M}_{\mathrm{prior}}$ is defined in (2). Hence, as long as prior (2) is considered, $\hat{\boldsymbol{h}}_{\mathrm{SS}}$ corresponds to the element of $\mathcal{H}$ leading to the best worst-case approximation error.

4 The multi-space approximation

In this section, we present a theoretical result supporting the performance of the “multi-space” decoder considered in [1, 6–8]. We state our main result in Theorem 2 and provide two examples of scenarios in which the multi-space decoder has provably better performance guarantees than the standard “single space” decoder (10). We then provide a graphical illustration of the geometry underlying the single-space and the multi-space problems. Finally, we elaborate on the differences and similarities between the results presented in this paper and those in [3].

4.1 Recovery guarantee for the multi-space decoder considered in [1, 6–8]

Before stating our result, we recall the definition of the multi-space decoder considered in [1, 6–8]:[5]

$$\begin{aligned} \text{Find }\ \hat{\boldsymbol{h}}_{\mathrm{MS}} \in \operatorname*{arg\,min}_{\boldsymbol{h}\in V_n} & \sum_{j=1}^{m} \left(y_j - \langle \boldsymbol{w}_j, \boldsymbol{h}\rangle\right)^2 \\ & \text{subject to } \operatorname{dist}(\boldsymbol{h}, V_k) \leq \hat{\epsilon}_k, \quad k = 0\ldots n. \end{aligned} \tag{16}$$

In the sequel, we assume without loss of generality that the vectors $\{\boldsymbol{w}_j\}_{j=1}^m$ are linearly independent (but not necessarily orthonormal). We also suppose that the subspaces $\{V_k\}_{k=0}^n$ are embedded, that is, obey (4).

[5] In this paper we assume that the constraints are defined $\forall k \in \{0\ldots n\}$. All the derivations presented hereafter may nevertheless be easily extended to the case where the constraints in (16) are only available for *some* $k \in \{0\ldots n\}$.

We note that (16) can be seen as an extension of the single-space decoder (10) discussed in Section 3. More specifically, if one removes the constraints in (16), the problem becomes

$$\text{Find } \hat{\boldsymbol{h}} \in \operatorname*{arg\,min}_{\boldsymbol{h}\in V_n} \sum_{j=1}^{m} (y_j - \langle \boldsymbol{w}_j, \boldsymbol{h}\rangle)^2. \tag{17}$$

It is then quite easy to see that the unique solution of this problem is given by (10) as soon as $m = n$ and $\sigma_{\min} > 0$. On the other hand, we note that the constraints in (16) define a set of feasible points of the form (3). Hence, if the subspaces $\{V_k\}_{k=0}^n$ and widths $\{\hat{\epsilon}_k\}_{k=0}^n$ are properly chosen, these constraints add some valuable information about the position of the sought solution $\boldsymbol{h}^\star$ in $\mathcal{H}$. We may thus expect the multi-space decoder (16) to lead to enhanced performance in some specific situations. In this section, we provide a mathematical support to this intuition.

We remind the reader that we assume that $\boldsymbol{h}^\star \in \mathcal{M}_{\text{prior}}$. When multi-space prior (3) is considered, this entails that

$$\operatorname{dist}(\boldsymbol{h}^\star, V_k) \le \hat{\epsilon}_k \text{ for all } k = 0 \dots n, \tag{18}$$

that is the target solution $\boldsymbol{h}^\star$ is a feasible point of the optimization problem (16). Under this assumption, we provide a mathematical characterization of the performance achievable by the multi-space decoder (16). More specifically, we derive an upper bound on the approximation error $\|\hat{\boldsymbol{h}}_{\text{MS}} - \boldsymbol{h}^\star\|$ depending on the distance between $\boldsymbol{h}^\star$ and the different approximation subspaces $\{V_k\}_{k=0}^n$. Our result is presented in Theorem 2 below.

In order to state our result we need to introduce the following quantities. We first define the short-hand notations

$$\epsilon_k = \operatorname{dist}(\boldsymbol{h}^\star, V_k) \text{ for all } k = 0 \dots n, \tag{19}$$

and

$$\gamma = \sup_{\boldsymbol{h}\in V_n^\perp, \|\boldsymbol{h}\|=1} \left(\sum_{j=1}^{m} \langle \boldsymbol{w}_j, \boldsymbol{h}\rangle^2 \right)^{\frac{1}{2}}, \tag{20}$$

where $V_n^\perp$ is the orthogonal complement of V_n in $\mathcal{H}$. We let $\{\boldsymbol{v}_j\}_{j=1}^n$ be an orthonormal basis of V_n such that

$$V_k = \operatorname{span}\left(\{\boldsymbol{v}_j\}_{j=1}^k\right). \tag{21}$$

We note that such a basis always exists since we assume that the sequence of subspaces $\{V_k\}_{k=0}^n$ obeys (4).

We define the Gram matrix $\mathbf{G}$ as

$$\mathbf{G} = [\langle \boldsymbol{w}_i, \boldsymbol{v}_j\rangle]_{i,j} \in \mathbb{R}^{m\times n}, \tag{22}$$

and let

$$\delta_j = \sum_{k=1}^{n} |x_{kj}|(\hat{\epsilon}_{k-1} + \epsilon_{k-1}) \qquad j = 1 \dots n, \tag{23}$$

where x_{kj} are the elements of matrix $\mathbf{X}$ appearing in the singular value decomposition of $\mathbf{G}$, that is $\mathbf{G} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}$, where $\mathbf{U} \in \mathbb{R}^{m\times m}$, $\mathbf{X} \in \mathbb{R}^{n\times n}$ are orthogonal matrices and $\boldsymbol{\Lambda} \in \mathbb{R}^{m\times n}$ is the diagonal matrix of singular values $\{\sigma_j\}_{j=1}^{\min(m,n)}$. In the sequel we will consider the extended set $\{\sigma_j\}_{j=1}^{n}$ by using the following convention: if $n > m$, we define $\sigma_j = 0$ for all $j > m$. Without loss of generality, we assume that the singular values $\{\sigma_j\}_{j=1}^{n}$ are sorted by decreasing order of magnitude.

Finally, we introduce the function

$$B\left(\{\delta_j\}_{j=1}^{n}, \tau\right) = \begin{cases} \left(\sum_{j=\ell+1}^{n} \delta_j^2 + \rho\, \delta_\ell^2 + \epsilon_n^2\right)^{\frac{1}{2}} & \text{if } \sum_{j=1}^{n} \sigma_j^2 \delta_j^2 \geq \tau^2, \\ \left(\sum_{j=1}^{n} \delta_j^2 + \epsilon_n^2\right)^{\frac{1}{2}} & \text{otherwise,} \end{cases} \tag{24}$$

where ℓ is the largest integer such that

$$\sum_{j=\ell}^{n} \sigma_j^2 \delta_j^2 \geq \tau^2, \tag{25}$$

and $\rho \in [0, 1]$ is defined as

$$\rho \sigma_\ell^2 \delta_\ell^2 + \sum_{j=\ell+1}^{n} \sigma_j^2 \delta_j^2 = \tau^2. \tag{26}$$

Using these notations, our result reads:

Theorem 2 *Assume $\boldsymbol{h}^\star$ verifies* (18) *and let $y_j = \langle \boldsymbol{w}_j, \boldsymbol{h}^\star \rangle$ for $j = 1 \dots m$. Then any solution $\hat{\boldsymbol{h}}_{\mathrm{MS}}$ of* (16) *verifies*

$$\left\|\boldsymbol{h}^\star - \hat{\boldsymbol{h}}_{\mathrm{MS}}\right\| \leq B\left(\{\delta_j\}_{j=1}^{n}, 2\gamma\epsilon_n\right). \tag{27}$$

Moreover, if $\sigma_n > 0$, (16) *admits a unique solution.*

The proof of Theorem 2 is presented in Section 5. We can make the following remarks about the result presented in Theorem 2. First, we note that all the quantities appearing in (24)-(26) are easily computable so that the upper bound $B(\{\delta_j\}_{j=1}^{n}, 2\gamma\epsilon_n)$ can be straightforwardly evaluated numerically.[6] We also notice that the performance guarantee stated in Theorem 2 is valid for any values of m and n. In particular, if $\hat{\epsilon}_k < \infty$ for all $k = 0 \dots n$, we always

[6] If $\{\epsilon_k = \mathrm{dist}(\boldsymbol{h}^\star, V_k)\}_{k=1}^{n}$ is known, the result provides an upper bound on the reconstruction error for a particular $\boldsymbol{h}^\star$. If not, setting $\epsilon_k = \hat{\epsilon}_k$ provides an "a priori" upper bound on the reconstruction error of any $\boldsymbol{h}^\star$ verifying (18).

have $B(\{\delta_j\}_{j=1}^n, 2\gamma\epsilon_n) < \infty$ since all the terms in the right-hand side of (24) are finite. This means that the multi-space decoder (16) always leads to a bounded error even if $m < n$. This is in good accordance with the fact that both $\boldsymbol{h}^\star$ and $\hat{\boldsymbol{h}}_{\text{MS}}$ belong to the bounded set $\mathcal{M}_{\text{prior}} \cap \mathcal{P}_{\mathbf{y}}$. We note however that when $m < n$, we have $\sigma_n = 0$ and the solution of (16) is usually no longer unique.

Finally, let us point out that the upper bound (27) is not sharp. This can for example be seen when $m = n$, $\hat{\epsilon}_k \to \infty$ for $k = 0 \ldots n$, and $\{\boldsymbol{w}_j\}_{j=1}^m$ is an orthonormal basis. In this case, problem (16) reduces[7] to (17) which, in turn, is equivalent to the single-space problem (10) as soon as $\sigma_n > 0$.[8] Considering the latter setup, we thus have $\hat{\boldsymbol{h}}_{\text{MS}} = \hat{\boldsymbol{h}}_{\text{SS}}$ and a sharp upper bound on the performance of the multi-space decoder is given by (13) in Theorem 1:

$$\left\| \boldsymbol{h}^\star - \hat{\boldsymbol{h}}_{\text{SS}} \right\| \leq \sigma_n^{-1}\, \epsilon_n. \tag{28}$$

On the other hand, particularizing (27) to this setup, we easily obtain that

$$\left\| \boldsymbol{h}^\star - \hat{\boldsymbol{h}}_{\text{SS}} \right\| \leq \left(4\sigma_n^{-2} + 1\right)^{\frac{1}{2}} \epsilon_n, \tag{29}$$

which is slightly more pessimistic than (28). In the next section, we nevertheless emphasize that the result stated in Theorem 2 is sufficiently accurate to show the superiority of the multi-space decoder over its single-space counterpart in some specific setups.

4.2 Comparison with the single-space approximation

We now turn our attention to the comparison between the performance achievable by the standard single-space decoder (10) and its multi-space version (16). More specifically, we show that the recovery guarantees obtained in the multi-space setup may be much more favorable than in the single-space setup in some specific settings. In order to allow a comparison between the single and multi-space decoders, we consider the case where $m = n$ and assume that $\{\boldsymbol{w}_j\}_{j=1}^m$ is an orthonormal basis. We note that, in such a case, we have $\sigma_{\min} = \sigma_n$ and $\gamma \leq 1$. The specific settings considered hereafter are inspired from [3] and are described in Examples 1 and 2. They correspond to different choices of matrix $\mathbf{X}$ appearing in the singular value decomposition of the Gram matrix $\mathbf{G}$ defined in (22). Since the latter matrix directly depends on the bases $\{\boldsymbol{w}_j\}_{j=1}^m$ and $\{\boldsymbol{v}_j\}_{j=1}^n$, these examples thus correspond to some particular choices of the observation and approximation subspaces W_m and $\{V_k\}_{k=0}^n$.

[7] In particular, all the constraints in (16) become inactive.

[8] We note that $\sigma_{\min} = \sigma_n$ because $\{\boldsymbol{w}_j\}_{j=1}^m$ is an orthonormal basis.

Example 1 We first assume that $\mathbf{X} = \mathbf{I}_n$ in the singular-value decomposition of $\mathbf{G}$. We set $\hat{\epsilon}_j = \epsilon_j$ and assume that

$$\epsilon_j = \begin{cases} 1 & j = 0 \dots n-3, \\ \epsilon^{\frac{1}{2}} & j = n-2, n-1, \\ \epsilon & j = n, \end{cases} \tag{30}$$

for some $\epsilon \ll 1$. Moreover, we let

$$\sigma_j = \begin{cases} 1 & j = 1 \dots n-3, \\ \epsilon^{\frac{1}{2}} & j = n-2, n-1, \\ \epsilon & j = n. \end{cases} \tag{31}$$

In this setup, the upper bound (13) of Theorem 1 becomes:

$$\left\| \hat{\boldsymbol{h}}_{\mathrm{SS}} - \boldsymbol{h}^\star \right\| \leq \sigma_n^{-1} \operatorname{dist}(\boldsymbol{h}^\star, V_n) = \epsilon^{-1}\epsilon = 1. \tag{32}$$

We note that this is sharp. In particular, there exists some $\boldsymbol{h}^\star \in \mathcal{H}$ such that $\operatorname{dist}(\boldsymbol{h}^\star, V_n) = \epsilon_n$ and the equality holds in (32) (see second part of Theorem 1).

On the other hand, because $\mathbf{X} = \mathbf{I}_n$ and $\hat{\epsilon}_j = \epsilon_j$, we have

$$\delta_j = \hat{\epsilon}_{j-1} + \epsilon_{j-1} = 2\epsilon_{j-1}. \tag{33}$$

The index ℓ appearing in the definition of the function B in (24) is then smaller or equal to $n-1$ since

$$\begin{aligned} \sigma_n^2 \delta_n^2 &= \sigma_n^2 (2\epsilon_{n-1})^2 = 4\epsilon^3 \ll 4\epsilon^2, \\ \sigma_{n-1}^2 \delta_{n-1}^2 &= \sigma_{n-1}^2 (2\epsilon_{n-2})^2 = 4\epsilon^2, \end{aligned}$$

and thus

$$\sigma_{n-1}^2 \delta_{n-1}^2 + \sigma_n^2 \delta_n^2 \geq 4\epsilon^2 \geq 4\gamma^2\epsilon^2 \tag{34}$$

since $\gamma \leq 1$. The upper bound in Theorem 2 becomes

$$\begin{aligned} \left\| \boldsymbol{h}^\star - \hat{\boldsymbol{h}}_{\mathrm{MS}} \right\| &\leq \left(\delta_{n-1}^2 + \delta_n^2 + \epsilon_n^2 \right)^{\frac{1}{2}}, \\ &= \left(4\epsilon + 4\epsilon + \epsilon^2 \right)^{\frac{1}{2}}, \\ &\leq 3\epsilon^{\frac{1}{2}}. \end{aligned} \tag{35}$$

Hence the bound in the multi-space setup (35) can be arbitrarily small as compared to (32) when $\epsilon \to 0$.

□

Example 2 We now consider $\mathbf{X} = n^{-\frac{1}{2}}\,\mathbf{1}_{n\times n}$ where $\mathbf{1}_{n\times n}$ is an $n \times n$ matrix of 1's. We set $\hat{\epsilon}_j = \epsilon_j$ and assume that

$$\epsilon_j = \begin{cases} \frac{1}{2} & j = 0, \\ \frac{1}{2(n-1)} & j = 1 \dots n-1, \\ \epsilon & j = n, \end{cases} \tag{36}$$

for some $\epsilon \ll n^{-1}$ (Note that we must have: $\epsilon \leq \frac{1}{2(n-1)}$ by definition). Moreover, we let

$$\sigma_j = \begin{cases} \sigma & j = 1 \dots n-1, \\ \epsilon^2 & j = n, \end{cases} \tag{37}$$

for some $1 \geq \sigma > \epsilon$ whose value will be specified below.

With these choices, the upper bound (13) of Theorem 1 becomes:

$$\left\|\hat{\boldsymbol{h}}_{\mathrm{SS}} - \boldsymbol{h}^\star\right\| \leq \sigma_n^{-1}\ \mathrm{dist}(\boldsymbol{h}^\star, V_n) = \epsilon^{-2}\epsilon = \epsilon^{-1}. \tag{38}$$

On the other hand, we have

$$\begin{aligned} \delta_j &= \sum_{k=1}^{n} |x_{kj}|(\hat{\epsilon}_{k-1} + \epsilon_{k-1}), \\ &= 2n^{-\frac{1}{2}} \sum_{k=1}^{n} \epsilon_{k-1}, \\ &= 2n^{-\frac{1}{2}}. \end{aligned} \tag{39}$$

By choosing σ such that (we remind the reader that $\sigma_{n-1} = \sigma$ by definition (37))

$$\sigma_{n-1}^2\delta_{n-1}^2 + \sigma_n^2\delta_n^2 = 4\epsilon^2, \tag{40}$$

we obtain that index ℓ appearing in the definition of the function B in (24) is smaller or equal to $n-1$ since $\gamma \leq 1$. The upper bound in Theorem 2 then reads

$$\begin{aligned} \left\|\boldsymbol{h}^\star - \hat{\boldsymbol{h}}_{\mathrm{MS}}\right\| &\leq \left(\delta_{n-1}^2 + \delta_n^2 + \epsilon_n^2\right)^{\frac{1}{2}}, \\ &= \left(4n^{-1} + 4n^{-1} + \epsilon^2\right)^{\frac{1}{2}}, \\ &\leq 3n^{-\frac{1}{2}}, \end{aligned} \tag{41}$$

where the last inequality follows from our initial assumption $\epsilon \ll n^{-1}$.

□

We conclude this section by providing a graphical representation of the geometry of problem (16). Fig. 1 gives an illustration of the feasible set and the iso-contours of the cost function appearing in (16); these quantities are plotted in the plane V_n for $m = n = 2$.

If $\sigma_n > 0$, it can be seen (see Appendix A) that the cost function $f(\boldsymbol{h}) \triangleq \sum_{j=1}^{n}(y_j - \langle \boldsymbol{w}_j, \boldsymbol{h}\rangle)^2$ can also be rewritten for any $\boldsymbol{h} \in V_n$ as

$$f(\boldsymbol{h}) = \sum_{j=1}^{n} \sigma_j^2 \Big(\Big\langle \boldsymbol{v}_j^*, \hat{\boldsymbol{h}}_{\mathrm{SS}} \Big\rangle - \langle \boldsymbol{v}_j^*, \boldsymbol{h} \rangle \Big)^2, \tag{42}$$

where

$$\boldsymbol{v}_j^* = \sum_{i=1}^{n} x_{ij} \boldsymbol{v}_i. \tag{43}$$

Here, the elements x_{ij}'s correspond to the components of the orthonormal matrix $\mathbf{X}$ appearing in the singular value decomposition of $\mathbf{G}$. From (42), we thus see that the iso-contours of $f(\boldsymbol{h})$ in V_n correspond to n-dimensional ellipsoids with center equal to $\hat{\boldsymbol{h}}_{\mathrm{SS}}$ and principal axes equal to $\{\boldsymbol{v}_j^*\}_{j=1}^{n}$. Moreover, the elongation of the ellipsoids along each axis $\boldsymbol{v}_j^*$ is inversely proportional to the singular value σ_j.

From a geometric point of view, bad recovery guarantees in the single-space setup corresponds to situations where the "iso-contour" ellipsoids is very elongated along (at least) one direction (*i.e.*, σ_n is small): the center of the ellipsoid $\hat{\boldsymbol{h}}_{\mathrm{SS}}$ may then be quite distant from the optimal orthogonal projection $P_{V_n}(\boldsymbol{h}^\star)$.

The prior information used in the multi-space decoder (16) may provide a solution to this problem by constraining $\hat{\boldsymbol{h}}_{\mathrm{MS}}$ to belong to some prespecified feasible set. Fig. 1 gives an illustration of such a situation. The feasible set defined by the constraints in (16) is represented by the gray shaded area. In the simplified setup considered here, it corresponds to the intersection of two sets: the constraint associated to V_0 imposes $\hat{\boldsymbol{h}}_{\mathrm{MS}}$ to belong to a ball of radius $\hat{\epsilon}_0$; the constraint corresponding to V_1 requires that $\hat{\boldsymbol{h}}_{\mathrm{MS}}$ does not deviate from the line passing through $\boldsymbol{v}_1$ by more than $\hat{\epsilon}_1$. The multi-space estimate $\hat{\boldsymbol{h}}_{\mathrm{MS}}$ then corresponds to the element of the feasible set leading to the smallest value of the cost function. We see in Fig. 1 that constraining the estimate $\hat{\boldsymbol{h}}_{\mathrm{MS}}$ to belong to the feasible set prevents it from deviating too far from $P_{V_n}(\boldsymbol{h}^\star)$. In particular, in the simple example described in Fig. 1, the multi-space estimate leads to better approximation performance than its single-space counterpart.

The same type of conclusions can in fact be drawn in more general setups: the multi-space decoder (16) is able to enhance the performance of the single-space approach (10) as soon as the prior information used in (16) can compensate for large deviations of the "iso-contour" ellipsoids. The gain achievable in the multi-space setup thus depends on the relative configuration of the "iso-contour ellipsoids" and the feasible set. We note that the shape of the ellipsoids

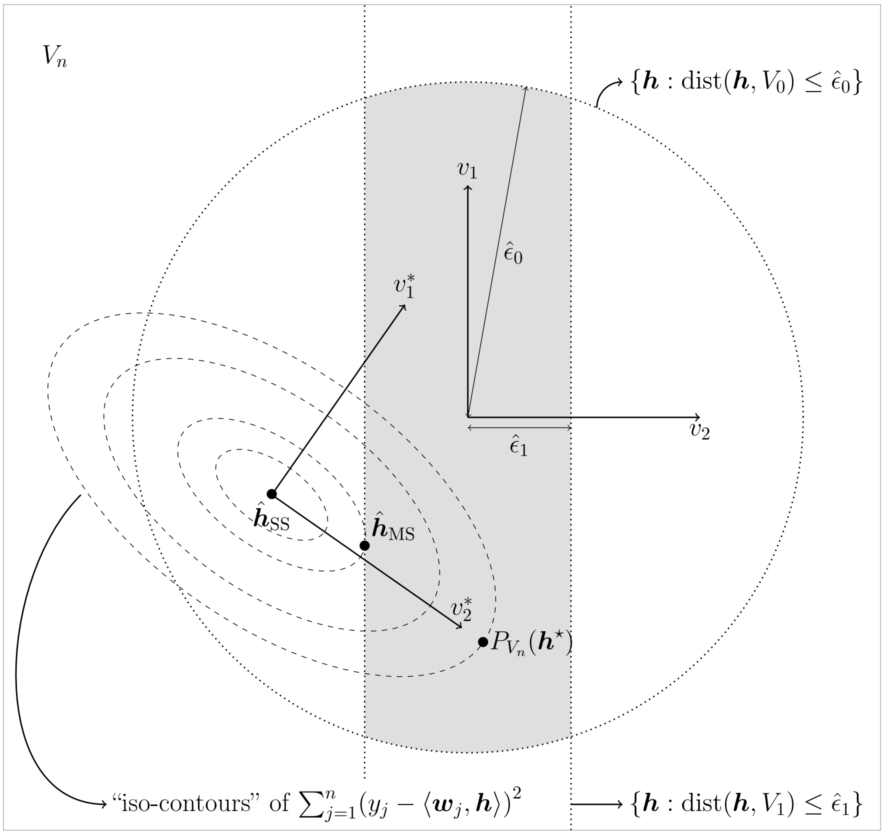

Fig. 1 Graphical representation of the geometry of the multi-space problem (16).

depends on the basis $\{\boldsymbol{v}_j^*\}_{j=1}^n$ and the singular value $\{\sigma_j\}_{j=1}^n$. Similarly, the feasible set is fully defined by the basis $\{\boldsymbol{v}_j\}_{j=1}^n$ and widths $\{\hat{\epsilon}_j\}_{j=0}^n$. Moreover, $\{\boldsymbol{v}_j\}_{j=1}^n$ and $\{\boldsymbol{v}_j^*\}_{j=1}^n$ only differ up to an orthogonal transformation $\mathbf{X}$, see (43). This explains why the parameters $\{\sigma_j\}_{j=1}^n$, $\{\hat{\epsilon}_j\}_{j=0}^n$ and $\{\delta_j\}_{j=1}^n$ (which depend on $\mathbf{X}$) play a crucial role in the characterization of the performance of the multi-space decoder in Theorem 2.

4.3 Comparison with the result by Binev *et al.* in [3]

In [3], the authors considered a different multi-space decoder:

$$\text{Find } \hat{\boldsymbol{h}}_{\mathrm{MSU}} \in \mathcal{M}_{\mathrm{prior}} \cap \mathcal{P}_{\mathbf{y}}, \tag{44}$$

where $\mathcal{M}_{\mathrm{prior}}$ is defined in (3). In this section, we elaborate on the connections and the differences between decoders (16) and (44).

Let us first mention that, although exploiting some "multi-space" prior, the two decoders differ in their definition. Whereas (16) addresses a variational problem where the estimate $\hat{\boldsymbol{h}}_{\mathrm{MS}}$ is the solution of a convex optimization problem, (44) focusses on a "feasility" problem, where one element in the intersection of $\mathcal{M}_{\mathrm{prior}}$ and $\mathcal{P}_{\mathbf{y}}$ must be found. The two decoders also differ in the constraints they impose on the solution: (44) enforces $\hat{\boldsymbol{h}}_{\mathrm{MSU}}$ to belong to the affine subspace $\mathcal{P}_{\mathbf{y}}$, whereas (16) imposes $\hat{\boldsymbol{h}}_{\mathrm{MS}} \in V_n$ but minimizes some distance (depending on $\{\boldsymbol{w}_j\}_{j=1}^m$) between $\hat{\boldsymbol{h}}_{\mathrm{MS}}$ and $\mathcal{P}_{\mathbf{y}}$. Although these choices may lead to some differences in the implementation of these decoders, we will restrict ourselves hereafter to a discussion of their performance.

In [3], the authors derived an upper bound on the approximation error achievable by the multi-space decoder (44). In order to state their result, we introduce the following notations. We first define

$$\tilde{\mathbf{G}} = [\langle \tilde{\boldsymbol{w}}_i, \boldsymbol{v}_j \rangle]_{i,j} \in \mathbb{R}^{m \times n}. \tag{45}$$

where $\{\tilde{\boldsymbol{w}}_i\}_{i=1}^m$ is an *orthonormal* basis of W_m. Moreover, we let

$$\tilde{\delta}_j = \sum_{k=1}^{n} |\tilde{x}_{kj}| \hat{\epsilon}_{k-1} \qquad j = 1 \dots n, \tag{46}$$

where $\tilde{x}_{kj}$ are the elements of matrix $\tilde{\mathbf{X}}$ appearing in the singular value decomposition of $\tilde{\mathbf{G}}$, that is $\tilde{\mathbf{G}} = \tilde{\mathbf{U}}\tilde{\boldsymbol{\Lambda}}\tilde{\mathbf{X}}^{\mathrm{T}}$ ($\tilde{\mathbf{U}} \in \mathbb{R}^{m\times m}$, $\tilde{\mathbf{X}} \in \mathbb{R}^{n \times n}$ are orthogonal matrices and $\tilde{\boldsymbol{\Lambda}} \in \mathbb{R}^{m\times n}$ is the diagonal matrix of singular values $\{\tilde{\sigma}_j\}_{j=1}^{\min(m,n)}$). We define the extended set $\{\tilde{\sigma}_j\}_{j=1}^n$ by using the following convention: if $n > m$, we set $\tilde{\sigma}_j = 0$ for all $j > m$. We assume that the singular values $\{\tilde{\sigma}_j\}_{j=1}^n$ are sorted by decreasing order of magnitude. The upper bound derived in [3] then reads:

Theorem 3 ([3, Section 3]) *Assume $\boldsymbol{h}^\star$ verifies* (18)*. Then any solution $\hat{\boldsymbol{h}}_{\mathrm{MSU}}$ of* (44) *verifies*

$$\left\| \boldsymbol{h}^\star - \hat{\boldsymbol{h}}_{\mathrm{MSU}} \right\| \leq 4B\left(\{\tilde{\delta}_j\}_{j=1}^n, \hat{\epsilon}_n\right). \tag{47}$$

We note that, due to the close connection between problems (16) and (44), the results stated in Theorems 2 and 3 are unsurprinsingly similar although involving different quantities and constants. On the one hand, the quantities appearing in Theorem 2 (*e.g.*, $\{\sigma_j\}_{j=1}^n$, $\{\delta_j\}_{j=1}^n$ or γ) depends on the basis $\{\boldsymbol{w}_j\}_{j=1}^m$ used to make the measurements $\{y_j\}_{j=1}^m$. On the other hand, $\{\tilde{\sigma}_j\}_{j=1}^n$ and $\{\tilde{\delta}_j\}_{j=1}^n$ used in Theorem 3 are only functions of the subspace W_m but not specifically on the generating vectors $\{\boldsymbol{w}_j\}_{j=1}^m$.[9] This is in good accordance with the definitions of the decoders in (16) and (44) as discussed at the beginning of this section.

[9] Indeed, we note that the singular values $\{\tilde{\sigma}_j\}_{j=1}^n$ and matrix $\tilde{\mathbf{X}}$ only depend on the subspace W_m but not on the orthonormal basis $\{\tilde{\boldsymbol{w}}_i\}_{i=1}^m$ used to define it.

Despite of their differences, the upper bounds stated in Theorems 2 and 3 can be compared when $\{\boldsymbol{w}_j\}_{j=1}^m$ is an orthonormal basis and $\hat{\epsilon}_k = \epsilon_k \ \forall k$. In this case, simple calculations show that

$$\sigma_j = \tilde{\sigma}_j, \ \ \delta_j = 2\tilde{\tilde{\delta}}_j, \ \ \gamma \leq 1, \tag{48}$$

and (27) particularizes to

$$\left\|\boldsymbol{h}^\star - \hat{\boldsymbol{h}}_{\mathrm{MS}}\right\| \leq 2B\left(\{\tilde{\delta}_j\}_{j=1}^n, \hat{\epsilon}_n\right). \tag{49}$$

Comparing (49) to (47), we then see that our result in Theorem 2 is more favorable than Theorem 3 by a factor 2. However, no conclusion on the *relative* performance of the multi-space decoders can be drawn from Theorems 2 and 3 since the bounds presented in these results are generally not sharp. In fact, we emphasize herefafter that no general ordering of the performance of decoders (16) and (44) can be made from a "worst-case" perspective, *i.e.*, by considering the following figure of merit

$$E(\hat{\boldsymbol{h}}) \triangleq \sup_{\boldsymbol{h}\in\mathcal{M}_{\mathrm{prior}}\cap\mathcal{P}_{\mathbf{y}}} \left\|\boldsymbol{h} - \hat{\boldsymbol{h}}\right\|, \tag{50}$$

where $\mathcal{M}_{\mathrm{prior}}$ is defined in (3). In particular, we provide two examples of scenarios where either decoder (16) outperforms (44) or vice-versa.

First, we note that since $\mathcal{M}_{\mathrm{prior}} \cap \mathcal{P}_{\mathbf{y}}$ is a bounded, closed, convex set, $E(\hat{\boldsymbol{h}})$ admits a unique minimizer which necessarily belongs to $\mathcal{M}_{\mathrm{prior}} \cap \mathcal{P}_{\mathbf{y}}$, see [3, Lemma 2.1 & Remark 2.4]. We denote this minimizer $\hat{\boldsymbol{h}}_{\mathrm{WC}}$ in what follows. From this remark, the scenario where (44) achieves better performance than (16) is as follows: since $\hat{\boldsymbol{h}}_{\mathrm{MSU}} \in \mathcal{M}_{\mathrm{prior}} \cap \mathcal{P}_{\mathbf{y}}$, it may happen that $\hat{\boldsymbol{h}}_{\mathrm{MSU}}$ is (luckily) equal to $\hat{\boldsymbol{h}}_{\mathrm{WC}}$. In such a case, (44) thus attains the best-achievable worst-case performance. On the other hand, if $\mathcal{P}_{\mathbf{y}} \cap V_n = \emptyset$, we have necessarily that $\hat{\boldsymbol{h}}_{\mathrm{MS}} \notin \mathcal{M}_{\mathrm{prior}} \cap \mathcal{P}_{\mathbf{y}}$ and, from the unicity of $\hat{\boldsymbol{h}}_{\mathrm{WC}}$, $E(\hat{\boldsymbol{h}}_{\mathrm{MSU}}) < E(\hat{\boldsymbol{h}}_{\mathrm{MS}})$.

A scenario where the reverse inequality holds is as follows. Assume that $\mathcal{P}_{\mathbf{y}} \cap V_n \neq \emptyset$. In this case, $\hat{\boldsymbol{h}}_{\mathrm{WC}} \in V_n$ as soon as $\mathcal{M}_{\mathrm{prior}} \cap \mathcal{P}_{\mathbf{y}}$ is symmetric around some point in V_n (see [3, Remark 2.2]). Moreover, $\hat{\boldsymbol{h}}_{\mathrm{WC}}$ is a solution[10] of the constrained problem (16) since it satisfies the problem's constraints and leads to a value of the cost function equal to zero. In this case, (16) thus attains the optimal worst-case performance. On the other hand, since (44) may output any point of $\mathcal{M}_{\mathrm{prior}} \cap \mathcal{P}_{\mathbf{y}}$, we do not have necessarily that $\hat{\boldsymbol{h}}_{\mathrm{MSU}} = \hat{\boldsymbol{h}}_{\mathrm{WC}}$. In the latter case, we thus have $E(\hat{\boldsymbol{h}}_{\mathrm{MS}}) < E(\hat{\boldsymbol{h}}_{\mathrm{MSU}})$.

5 Proof of Theorem 2

In this section, we provide a proof of the result stated in Theorem 2. We first note that problem (16) is equivalent to finding the minimum of a quadratic

[10] It is in fact the unique solution as soon as $n \leq m$ and the vectors $\{\boldsymbol{w}_j\}_{j=1}^m$ are linearly independent.

function over a closed bounded subset of V_n. A minimizer thus always exists. Moreover, the unicity of the minimizer stated at the end of Theorem 2 follows from the strict convexity of the cost function over V_n when $\sigma_n > 0$.

In the rest of this section, we thus mainly focus on the derivation of the upper bound (27). Our proof is based on the following steps. First, since $\hat{\boldsymbol{h}}_{\mathrm{MS}} \in V_n$, we have that

$$\begin{aligned}\left\|\boldsymbol{h}^\star - \hat{\boldsymbol{h}}_{\mathrm{MS}}\right\|^2 &= \left\|P_{V_n}(\boldsymbol{h}^\star) - \hat{\boldsymbol{h}}_{\mathrm{MS}}\right\|^2 + \left\|P_{V_n}^{\perp}(\boldsymbol{h}^\star)\right\|^2, \\ &= \left\|P_{V_n}(\boldsymbol{h}^\star) - \hat{\boldsymbol{h}}_{\mathrm{MS}}\right\|^2 + \epsilon_n^2, \end{aligned} \tag{51}$$

where $P_{V_n}(\cdot)$ (resp. $P_{V_n}^{\perp}(\cdot)$) denotes the orthogonal projector onto V_n (resp. $V_n^{\perp}$). We then derive an upper bound on $\|P_{V_n}(\boldsymbol{h}^\star) - \hat{\boldsymbol{h}}_{\mathrm{MS}}\|^2$ as follows:

- We identify a set $\mathcal{D}$ such that $P_{V_n}(\boldsymbol{h}^\star) - \hat{\boldsymbol{h}}_{\mathrm{MS}} \in \mathcal{D}$ in Section 5.1. This implies in particular that $\|P_{V_n}(\boldsymbol{h}^\star) - \hat{\boldsymbol{h}}_{\mathrm{MS}}\|^2 \leq \sup_{\boldsymbol{d}\in\mathcal{D}}\|\boldsymbol{d}\|^2$.
- We derive the analytical expression of $\sup_{\boldsymbol{d}\in\mathcal{D}}\|\boldsymbol{d}\|^2$ as a function of the parameters $\{\epsilon_k\}_{k=0}^n$, $\{\hat{\epsilon}_k\}_{k=0}^n$ and $\{\sigma_k\}_{k=1}^n$ in Section 5.2.

Combining these results, we obtain (27).

5.1 Definition of $\mathcal{D}$

We express $\mathcal{D}$ as the intersection of two sets $\mathcal{D}_1$ and $\mathcal{D}_2$ that we define in Sections 5.1.2 and 5.1.3 respectively. In order to properly define these quantities, we introduce some particular bases for V_n and W_m in Section 5.1.1.

5.1.1 Some particular bases for V_n and W_m

Let

$$\mathbf{G} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}} \tag{52}$$

be the singular value decomposition of the Gram matrix defined in (22), where $\mathbf{U} \in \mathbb{R}^{m\times m}$ and $\mathbf{X} \in \mathbb{R}^{n\times n}$ are orthonormal matrices and $\boldsymbol{\Lambda} \in \mathbb{R}^{m\times n}$ is the diagonal matrix of singular values. We denote by $\{\sigma_j\}_{j=1}^n$ the set of singular values of $\mathbf{G}$ sorted in their decreasing order of magnitude. We remind the reader that if $m < n$, we adopt the convention $\sigma_j = 0\ \forall j > m$.

We define the following bases for V_n and W_m:

$$\boldsymbol{v}_j^* = \sum_{i=1}^n x_{ij}\boldsymbol{v}_i, \tag{53}$$

$$\boldsymbol{w}_j^* = \sum_{i=1}^m u_{ij}\boldsymbol{w}_i, \tag{54}$$

where $\mathbf{U} \in \mathbb{R}^{m\times m}$ and $\mathbf{X} \in \mathbb{R}^{n\times n}$ are the orthonormal matrices appearing in (52). We note that $\{\boldsymbol{v}_j^*\}_{j=1}^n$ is an orthonormal basis whereas $\{\boldsymbol{w}_j^*\}_{j=1}^m$ is not necessarily orthonormal. By definition, $\{\boldsymbol{v}_j^*\}_{j=1}^n$ and $\{\boldsymbol{w}_j^*\}_{j=1}^m$ enjoy the following desirable property:

$$\langle \boldsymbol{w}_i^*, \boldsymbol{v}_j^* \rangle = \begin{cases} \sigma_j & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases} \tag{55}$$

5.1.2 Definition of $\mathcal{D}_1$

Let us define $\mathcal{D}_1$ as

$$\mathcal{D}_1 = \left\{ \boldsymbol{d} = \sum_{j=1}^n \beta_j \boldsymbol{v}_j^* : \sum_{j=1}^n \sigma_j^2 \beta_j^2 \leq 4\gamma^2 \epsilon_n^2 \right\}, \tag{56}$$

where γ is defined in (20). We show hereafter that $P_{V_n}(\boldsymbol{h}^\star) - \hat{\boldsymbol{h}}_{\mathrm{MS}} \in \mathcal{D}_1$.

Let us first consider the intermediate set

$$\mathcal{S} = \{ \boldsymbol{h} : f(\boldsymbol{h}) \leq \gamma^2 \epsilon_n^2 \}, \tag{57}$$

where $f(\boldsymbol{h}) \triangleq \sum_{j=1}^m (y_j - \langle \boldsymbol{w}_j, \boldsymbol{h} \rangle)^2$ is the cost function appearing in the variational formulation of multi-space decoder (16).

Clearly $P_{V_n}(\boldsymbol{h}^\star) \in \mathcal{S}$ because

$$\begin{aligned} f(P_{V_n}(\boldsymbol{h}^\star)) &= \sum_{j=1}^m \left(y_j - \langle \boldsymbol{w}_j, P_{V_n}(\boldsymbol{h}^\star) \rangle \right)^2 \\ &= \sum_{j=1}^m \left(\langle \boldsymbol{w}_j, \boldsymbol{h}^\star \rangle - \langle \boldsymbol{w}_j, P_{V_n}(\boldsymbol{h}^\star) \rangle \right)^2 \\ &= \sum_{j=1}^m \left(\langle \boldsymbol{w}_j, P_{V_n}^\perp(\boldsymbol{h}^\star) \rangle \right)^2 \\ &\leq \gamma^2 \left\| P_{V_n}^\perp(\boldsymbol{h}^\star) \right\|^2 \\ &\leq \gamma^2 \epsilon_n^2. \end{aligned} \tag{58}$$

Moreover, we also have $\hat{\boldsymbol{h}}_{\mathrm{MS}} \in \mathcal{S}$. This can be seen from the following arguments. First, $P_{V_n}(\boldsymbol{h}^\star)$ is a feasible point for problem (16), that is

$$\mathrm{dist}(P_{V_n}(\boldsymbol{h}^\star), V_k) \leq \hat{\epsilon}_k \text{ for } k = 0 \dots n. \tag{59}$$

Indeed, rewriting $\boldsymbol{h}^\star$ as

$$\boldsymbol{h}^\star = \sum_{j=1}^n \langle \boldsymbol{v}_j, \boldsymbol{h}^\star \rangle \boldsymbol{v}_j + \boldsymbol{z}, \tag{60}$$

where $\boldsymbol{z} \in V_n^{\perp}$, we have

$$
\begin{aligned}
\hat{\epsilon}_k &\geq \operatorname{dist}(\boldsymbol{h}^\star, V_k) \\
&= \left\| P_{V_k}^{\perp}(\boldsymbol{h}^\star) \right\| \\
&= \left\| \sum_{j=k+1}^{n} \langle \boldsymbol{v}_j, \boldsymbol{h}^\star \rangle \boldsymbol{v}_j + \boldsymbol{z} \right\| \\
&= \sqrt{\left\| \sum_{j=k+1}^{n} \langle \boldsymbol{v}_j, \boldsymbol{h}^\star \rangle \boldsymbol{v}_j \right\|^2 + \|\boldsymbol{z}\|^2} \\
&\geq \left\| \sum_{j=k+1}^{n} \langle \boldsymbol{v}_j, \boldsymbol{h}^\star \rangle \boldsymbol{v}_j \right\| \\
&= \left\| P_{V_k}^{\perp}\left(P_{V_n}(\boldsymbol{h}^\star)\right) \right\| \\
&= \operatorname{dist}(P_{V_n}(\boldsymbol{h}^\star), V_k).
\end{aligned} \tag{61}
$$

The first inequality follows from our initial assumption (18). The third equality is true because $\boldsymbol{z} \in V_n^{\perp}$. Now, since $\hat{\boldsymbol{h}}_{\mathrm{MS}}$ is a minimizer of $f(\boldsymbol{h})$ over the set of feasible points, we have $f(\hat{\boldsymbol{h}}_{\mathrm{MS}}) \leq f(P_{V_n}(\boldsymbol{h}^\star)) \leq \gamma^2 \epsilon_n^2$ and therefore $\hat{\boldsymbol{h}}_{\mathrm{MS}} \in \mathcal{S}$.

Finally, we show that $\hat{\boldsymbol{h}}_{\mathrm{MS}} \in \mathcal{S}$ and $P_{V_n}(\boldsymbol{h}^\star) \in \mathcal{S}$ implies $P_{V_n}(\boldsymbol{h}^\star) - \hat{\boldsymbol{h}}_{\mathrm{MS}} \in \mathcal{D}_1$. Let us first note that, if $\boldsymbol{h} \in V_n$, the cost function $f(\boldsymbol{h})$ can be rewritten as:

$$
\begin{aligned}
f(\boldsymbol{h}) &= \sum_{j=1}^{m} (\langle \boldsymbol{w}_j, \boldsymbol{h}^\star \rangle - \langle \boldsymbol{w}_j, \boldsymbol{h} \rangle)^2, \\
&= \sum_{j=1}^{m} \left(\left\langle \boldsymbol{w}_j^*, \boldsymbol{h}^\star \right\rangle - \left\langle \boldsymbol{w}_j^*, \boldsymbol{h} \right\rangle \right)^2, \\
&= \sum_{j=1}^{\min(m,n)} \left(\left\langle \boldsymbol{w}_j^*, \boldsymbol{h}^\star \right\rangle - \sigma_j \left\langle \boldsymbol{v}_j^*, \boldsymbol{h} \right\rangle \right)^2 + \sum_{j=n+1}^{m} \left\langle \boldsymbol{w}_j^*, \boldsymbol{h}^\star \right\rangle^2,
\end{aligned} \tag{62}
$$

where the second equality follows from the fact that $\{\boldsymbol{w}_j\}_{j=1}^m$ and $\{\boldsymbol{w}_j^*\}_{j=1}^m$ differ up to an orthonormal transformation; the last equality is a consequence of (55) and the fact that $\boldsymbol{h} \in V_n$ by hypothesis.

We note that, since $\{\boldsymbol{v}_j^*\}_{j=1}^n$ is an orthonormal basis of V_n, $P_{V_n}(\boldsymbol{h}^\star) - \hat{\boldsymbol{h}}_{\mathrm{MS}}$ can be written as $\sum_{j=1}^n \beta_j \boldsymbol{v}_j^*$ by setting $\beta_j = \left\langle \boldsymbol{v}_j^*, P_{V_n}(\boldsymbol{h}^\star) \right\rangle - \langle \boldsymbol{v}_j^*, \hat{\boldsymbol{h}}_{\mathrm{MS}} \rangle$.

Therefore, we have

$$\begin{aligned}
\sum_{j=1}^{n} \sigma_j^2 \beta_j^2 =& \sum_{j=1}^{n} \Big(\sigma_j \langle \boldsymbol{v}_j^*, P_{V_n}(\boldsymbol{h}^\star) \rangle - \sigma_j \Big\langle \boldsymbol{v}_j^*, \hat{\boldsymbol{h}}_{\text{MS}} \Big\rangle \Big)^2, \\
=& \sum_{j=1}^{n} \Big(\sigma_j \langle \boldsymbol{v}_j^*, P_{V_n}(\boldsymbol{h}^\star) \rangle - \langle \boldsymbol{w}_j^*, \boldsymbol{h}^\star \rangle - \sigma_j \Big\langle \boldsymbol{v}_j^*, \hat{\boldsymbol{h}}_{\text{MS}} \Big\rangle + \langle \boldsymbol{w}_j^*, \boldsymbol{h}^\star \rangle \Big)^2, \\
\leq& 2 \sum_{j=1}^{n} \big(\sigma_j \langle \boldsymbol{v}_j^*, P_{V_n}(\boldsymbol{h}^\star) \rangle - \langle \boldsymbol{w}_j^*, \boldsymbol{h}^\star \rangle \big)^2 + 2 \sum_{j=1}^{n} \Big(\sigma_j \Big\langle \boldsymbol{v}_j^*, \hat{\boldsymbol{h}}_{\text{MS}} \Big\rangle - \langle \boldsymbol{w}_j^*, \boldsymbol{h}^\star \rangle \Big)^2, \\
\leq& 2 f(P_{V_n}(\boldsymbol{h}^\star)) + 2 f(\hat{\boldsymbol{h}}_{\text{MS}}), \\
\leq& 4\gamma^2 \epsilon_n^2,
\end{aligned}$$

where the first inequality follows from the standard inequality $(a+b)^2 \leq 2(a^2+b^2)$, the second from (62), and the last one from the fact that $\hat{\boldsymbol{h}}_{\text{MS}} \in \mathcal{S}$ and $P_{V_n}(\boldsymbol{h}^\star) \in \mathcal{S}$.

5.1.3 Definition of $\mathcal{D}_2$

Let

$$\delta_j = \eta_j + \hat{\eta}_j, \tag{63}$$

where

$$\begin{aligned}
\eta_j &= \sum_{i=1}^{n} |x_{ij}| \epsilon_{i-1}, \\
\hat{\eta}_j &= \sum_{i=1}^{n} |x_{ij}| \hat{\epsilon}_{i-1},
\end{aligned} \tag{64}$$

and the x_{ij}'s are the elements of the matrix $\mathbf{X}$ appearing in the SVD decomposition (52). We define $\mathcal{D}_2$ as

$$\mathcal{D}_2 = \left\{ \boldsymbol{d} = \sum_{j=1}^{n} \beta_j \boldsymbol{v}_j^* : |\beta_j| \leq \delta_j \right\}. \tag{65}$$

We show hereafter that $P_{V_n}(\boldsymbol{h}^\star) - \hat{\boldsymbol{h}}_{\text{MS}} \in \mathcal{D}_2$.

We first note that if $\boldsymbol{h}$ is feasible for problem (16), we must have

$$\big| \langle \boldsymbol{v}_j^*, \boldsymbol{h} \rangle \big| \leq \hat{\eta}_j. \tag{66}$$

Indeed, if $\boldsymbol{h}$ is feasible, the constraint $\text{dist}(\boldsymbol{h}, V_k) \leq \hat{\epsilon}_k$ simply writes as

$$\sum_{j=k+1}^{n} \langle \boldsymbol{v}_j, \boldsymbol{h} \rangle^2 \leq \hat{\epsilon}_k^2.$$

In particular, this implies that

$$|\langle \boldsymbol{v}_{k+1}, \boldsymbol{h} \rangle| \leq \hat{\epsilon}_k.$$

Using the fact that

$$\boldsymbol{v}_j^* = \sum_{k=1}^{n} x_{kj} \boldsymbol{v}_k,$$

we obtain (66). In a similar way, we can find that

$$\left|\left\langle \boldsymbol{v}_j^*, P_{V_n}(\boldsymbol{h}^\star) \right\rangle\right| \leq \eta_j, \tag{67}$$

by using the fact that $\mathrm{dist}(P_{V_n}(\boldsymbol{h}^\star), V_k) \leq \epsilon_k$ from (61).

Let us now show that $P_{V_n}(\boldsymbol{h}^\star) - \hat{\boldsymbol{h}}_{\mathrm{MS}} \in \mathcal{D}_2$. Since $\{\boldsymbol{v}_j^*\}_{j=1}^n$ is an orthonormal basis of V_n, $P_{V_n}(\boldsymbol{h}^\star) - \hat{\boldsymbol{h}}_{\mathrm{MS}}$ can be written as $\sum_{j=1}^n \beta_j \boldsymbol{v}_j^*$ by setting $\beta_j = \langle \boldsymbol{v}_j^*, P_{V_n}(\boldsymbol{h}^\star) \rangle - \langle \boldsymbol{v}_j^*, \hat{\boldsymbol{h}}_{\mathrm{MS}} \rangle$. This leads to

$$\begin{aligned} |\beta_j| &= \left|\left\langle \boldsymbol{v}_j^*, P_{V_n}(\boldsymbol{h}^\star) \right\rangle - \left\langle \boldsymbol{v}_j^*, \hat{\boldsymbol{h}}_{\mathrm{MS}} \right\rangle\right|, \\ &\leq \left|\left\langle \boldsymbol{v}_j^*, P_{V_n}(\boldsymbol{h}^\star) \right\rangle\right| + \left|\left\langle \boldsymbol{v}_j^*, \hat{\boldsymbol{h}}_{\mathrm{MS}} \right\rangle\right|, \\ &\leq \hat{\eta}_j + \eta_j = \delta_j, \end{aligned}$$

where the last inequality follows from (66) and (67).

5.2 Expression of $\sup_{\boldsymbol{d} \in \mathcal{D}} \|\boldsymbol{d}\|^2$

We consider the following problem:

$$\sup_{\boldsymbol{d} \in \mathcal{D}} \|\boldsymbol{d}\|^2 = \sup_{\boldsymbol{\beta} \in \mathbb{R}^n} \|\boldsymbol{\beta}\|^2 \ \text{ subject to } \begin{cases} \sum_{j=1}^n \sigma_j^2 \beta_j^2 \leq 4\gamma^2 \epsilon_n^2 \\ |\beta_j| \leq \delta_j \end{cases}. \tag{68}$$

If $\sum_{j=1}^n \sigma_j^2 \delta_j^2 < 4\gamma^2 \epsilon_n^2$, the first constraint in (68) is always inactive and the solution simply reads

$$\sup_{\boldsymbol{d} \in \mathcal{D}} \|\boldsymbol{d}\|^2 = \sum_{j=1}^{n} \delta_j^2. \tag{69}$$

If $\sum_{j=1}^n \sigma_j^2 \delta_j^2 \geq 4\gamma^2 \epsilon_n^2$, the solution of (68) is given by

$$\sup_{\boldsymbol{d} \in \mathcal{D}} \|\boldsymbol{d}\|^2 = \sum_{j=\ell+1}^{n} \delta_j^2 + \rho\, \delta_\ell^2, \tag{70}$$

where ℓ is the largest integer such that

$$\sum_{j=\ell}^{n} \sigma_j^2 \delta_j^2 \geq 4\gamma^2 \epsilon_n^2, \tag{71}$$

and $\rho \in [0,1]$ is defined as

$$\rho\sigma_\ell^2\delta_\ell^2 + \sum_{j=\ell+1}^{n} \sigma_j^2\delta_j^2 = 4\gamma^2\epsilon_n^2. \tag{72}$$

This can be seen by verifying the optimality condition of problem (68). We note that problem (68) is the same (up to some constants) to the one considered in [3, Section 3.1]. The solution (70) is therefore similar, up to some different constants, to the one obtained in that paper.

6 Conclusions

In this paper, we provide a mathematical characterization of the performance of some particular decoder exploiting a set of approximation subspaces. This decoder was previously shown to lead to good empirical results in several contributions [1, 6–8], although no proof of its theoretical performance was provided in these works. Our result shows how the performance of this "multi-space" decoder is related to the parameters defining the approximation problem: the observation W_m and prior subspaces $\{V_k\}_{k=0}^n$, the quality of the prior information (*i.e.*, the widths $\{\hat{\epsilon}_k\}_{k=0}^n$) and the distance between the target solution $\boldsymbol{h}^\star$ and the approximation subspaces (*i.e.*, $\{\epsilon_k\}_{k=0}^n$). Based on this result, we show that the "multi-space" decoder can have provably better reconstruction guarantees than its standard ("single-space") counterpart in some situations.

A Proof of (42)

In this appendix, we show that the cost function $f(\boldsymbol{h}) \triangleq \sum_{j=1}^n (y_j - \langle \boldsymbol{w}_j, \boldsymbol{h}\rangle)^2$ can be rewritten as in (42) when $\boldsymbol{h} \in V_n$ and $\sigma_n > 0$.[11] First, using the definition of y_j, we have

$$f(\boldsymbol{h}) = \sum_{j=1}^{n} (\langle \boldsymbol{w}_j, \boldsymbol{h}^\star\rangle - \langle \boldsymbol{w}_j, \boldsymbol{h}\rangle)^2. \tag{73}$$

Moreover, using the particular bases introduced in Section 5.1.1, we obtain

$$\begin{aligned} f(\boldsymbol{h}) &= \sum_{j=1}^{n} \left(\langle \boldsymbol{w}_j^*, \boldsymbol{h}^\star\rangle - \langle \boldsymbol{w}_j^*, \boldsymbol{h}\rangle\right)^2, \\ &= \sum_{j=1}^{n} \left(\langle \boldsymbol{w}_j^*, \boldsymbol{h}^\star\rangle - \sigma_j\langle \boldsymbol{v}_j^*, \boldsymbol{h}\rangle\right)^2, \end{aligned} \tag{74}$$

where the first equality follows from the fact that $\{\boldsymbol{w}_j\}_{j=1}^n$ and $\{\boldsymbol{w}_j^*\}_{j=1}^n$ differ up to an orthogonal transformation; the second is a consequence of (55) and our hypothesis $\boldsymbol{h} \in V_n$.

Since $\hat{\boldsymbol{h}}_{\mathrm{SS}}$ corresponds to the minimum of $f(\boldsymbol{h})$ over V_n (see (17)), we simply have

$$\left\langle \boldsymbol{v}_j^*, \hat{\boldsymbol{h}}_{\mathrm{SS}} \right\rangle = \frac{\left\langle \boldsymbol{w}_j^*, \boldsymbol{h}^\star \right\rangle}{\sigma_j}, \tag{75}$$

if $\sigma_n > 0$. Hence, under this assumption, (74) can also be rewritten as in (42).

[11] We remind the reader that we assume $m = n$.

References

1. Argaud, J., Bouriquet, B., Gong, H., Maday, Y., Mula, O.: Stabilization of (G)EIM in presence of measurement noise: application to nuclear reactor physics. In: Spectral and High Order Methods for Partial Differential Equations ICOSAHOM 2016, pp. 133–145. Springer (2017)
2. Babuska, I.: Error-bounds for finite element method. Numerische Mathematik **16**, 322–333 (1970/71)
3. Binev, P., Cohen, A., Dahmen, W., DeVore, R., Petrova, G., Wojtaszczyk, P.: Data assimilation in reduced modeling. SIAM/ASA Journal on Uncertainty Quantification **5**(1), 1–29 (2017). DOI 10.1137/15M1025384. URL https://doi.org/10.1137/15M1025384
4. Chaturantabut, S., Sorensen, D.: Nonlinear model reduction via discrete empirical interpolation. SIAM J. Sci. Comput. **32**(5), 2737–2764 (2010). DOI 10.1137/090766498. URL http://dx.doi.org/10.1137/090766498
5. Everson, R., Sirovich, L.: Karhunen-Loève procedure for gappy data. J. Opt. Soc. Am. A **12**(8), 1657–1664 (1995). DOI 10.1364/josaa.12.001657. URL http://dx.doi.org/10.1364/josaa.12.001657
6. Fick, L., Maday, Y., Patera, A.T., Taddei, T.: A Reduced Basis Technique for Long-Time Unsteady Turbulent Flows. ArXiv e-prints (2017)
7. Herzet, C., Diallo, M., Héas, P.: Beyond petrov-galerkin projection by using multi-space prior. In: European Conference on Numerical Mathematics and Advanced Applications (Enumath'17). (https://hal.inria.fr/hal-02173637v1), Voss, Norway (2017)
8. Herzet, C., Diallo, M., Héas, P.: Beyond Petrov-Galerkin projection by using multi-space prior. In: Model Reduction of Parametrized Systems IV (MoRePaS'18). (https://hal.inria.fr/hal-01937876), Nantes, France (2018)
9. Maday, Y., Mula, O., Patera, A.T., Yano, M.: The generalized empirical interpolation method: Stability theory on hilbert spaces with an application to the stokes equation. Computer Methods in Applied Mechanics and Engineering **287**, 310–334 (2015). DOI 10.1016/j.cma.2015.01.018. URL http://www.sciencedirect.com/science/article/pii/S0045782515000389
10. Maday, Y., Mula, O., Turinici, G.: Convergence analysis of the generalized empirical interpolation method. SIAM Journal on Numerical Analysis **54**(3), 1713–1731 (2016)
11. Quarteroni, A., Manzoni, A., Negri, F.: Reduced Basis Methods for Partial Differential Equations, vol. 92. Springer International Publishing (2016). URL http://www.springer.com/us/book/9783319154305